CLINICAL INVESTIGATION

A review of randomization methods in clinical trials

Vance W Berger*[,1] & Olga Antsygina[2]
[1]National Cancer Institute & University of Maryland Baltimore County, Biometry Research Group, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA
[2]University of South Florida, College of Public Health, Department of Epidemiology & Biostatistics, 13201 Bruce B Downs, MDC56 Tampa, FL 33612, USA
*Author for correspondence:
Tel.: 240 276 7142
vb78c@nih.gov

In many respects, clinical trials can be seen as an art as well as a science, in that there is ample discretion for investigators to select research methods reflecting their own individual preferences. In fact, new research methods are developed on a fairly regular basis, not all of them improvements over existing methods. But the opposite trend also remains in effect, as researchers often follow established precedent, rather than thinking through the issues relevant to the current trial so as to come up with the research methods that are optimal in this case. These two forces pulling in opposite directions, individuality and inertia, compete in many aspects of clinical research, including the specific methods of randomization. New randomization methods are constantly proposed, while at the same time more and more researchers seem to be using the established standards of permuted blocks randomization or minimization (which, in its most extreme form, is not even true randomization at all). A comprehensive review of all randomization methods is beyond the scope of this work, but we will review these two established standards, as well as the newer (and vastly better) maximum tolerated imbalance procedures, including the big stick, Chen's procedure and the maximal procedure.

Keywords: allocation concealment • big stick • Chen's procedure • clinical trial • maximal procedure • minimization • MTI • permuted blocks • randomization methods • weak encryption

In many respects, clinical trials can be seen as an art as well as a science, in that there is ample discretion for investigators to select research methods reflecting their individual preferences. In fact, new research methods are developed on a fairly regular basis, not all of them improvements over existing methods. But the opposite trend also remains in effect, as researchers often follow the established precedent, rather than thinking through the issues relevant to the current trial so as to come up with the research methods that are optimal in this case [1]. So we have, simultaneously, too much conformity and too little conformity. These two forces pulling in opposite directions, individuality and inertia, compete in many aspects of clinical research, including the specific methods of randomization. New randomization methods are constantly proposed, while at the same time more and more researchers seem to be using the established standards of permuted blocked randomization (which provides only weak encryption) or minimization (which, in its most extreme form, is not even true randomization at all).

Because so many new randomization methods are proposed, a comprehensive review of all randomization methods is beyond the scope of this work, but we will review the aforementioned two established standards, permuted blocks and minimization, as well as the newer maximum tolerated imbalance (MTI) procedures, including the big stick [2], Chen's procedure [3] and the maximal procedure [4], with an eye toward comparing and contrasting them in terms of their ability to simultaneously control both chronological bias [5] and selection bias [6].

FUTURE SCIENCE part of fsg

In 'quasi-randomization' we shall distinguish true randomization from quasi-randomization. In the 'Minimization & adaptive procedures' section, we shall discuss minimization and other adaptive procedures. In the 'The permuted blocks design (PBD)' section, we offer a somber critique of the popular permuted blocks procedure. In 'MTI procedures' we discuss the much more appropriate MTI procedures (the big stick, Chen's procedure and the maximal procedure). In the 'Mixing & matching' section, we discuss combining various different randomization procedures so as to come up with something that is more robust than any one of the basic methods that was used to produce it. In the 'Executive summary', we offer a summary and our conclusions regarding how randomization should and should not be conducted.

Quasi-randomization

There are quite a few problems that plague clinical research, and these forces conspire to result in medical studies that, taken as a whole, are not reproducible. In fact, Altman [7] referred to medical research as a 'scandal', and Ioannidis [8] noted that most published research findings are false. One of the problems that has helped to get us there is a general tolerance for imprecise reporting. An overwhelming number of trial reports provide no information whatsoever regarding the precise methods of randomization used, beyond making the claim that the trial was randomized. Unfortunately, this claim is often false, and the misleading claim of randomization is almost never recognized as such, given this environment of trust without verifying.

Berger and Bears [9] distinguished quasi-randomization from true randomization, and noted how infrequently true randomization could be deduced from the descriptions that accompanied the claims of randomization. Far too often it turns out that alternation is used instead of randomization, and yet randomization is claimed anyway. In practice, we almost never know the difference, since authors are almost never held to any real standards in reporting what they did. And yet the distinction is a crucial one in terms of trial quality, reliability and validity.

In future sections we will highlight the difference between the MTI procedures, with their strong encryption and permuted blocks, with its weak encryption. But for now, we note that alternation offers no encryption at all. Indeed, in the case of an unmasked trial conducted with alternation instead of randomization, once we observe the identity of the first allocation, we will know with certainty all future ones as well. In other words, allocation concealment is impossible in this situation, and it remains impossible with alternation as long as there is even the chance of any allocations becoming unmasked. Therefore, alternation precludes the possibility of allocation concealment and it discredits results of trial.

In recognition of the distinction between alternation (and related procedures) and true randomization, some researchers refer to alternation as quasi-randomization. In fact, we too have done so in this paper, but this is done only to ensure that the procedures we condemn are recognized for what they are, since some researchers know them only as quasi-randomization, and never actually call them alternation. So there might otherwise be some danger that some readers would agree with our analysis, but then carry on using alternation thinking that they are using not alternation (which we condemn) but, rather, quasi-randomization, on which we remain silent. To be absolutely clear, then, we decided not to remain silent on the alter ego, and to instead call it out by name. This is not to suggest that we endorse or agree with the use of this highly misleading term.

The term 'quasi-randomization' suggests something just shy of true randomization but close enough that for all practical purposes we may safely ignore the technical distinction. It is all semantics anyway. But it is not all semantics. The distinction is real, and carries massive ramifications for the validity of the trial (or lack thereof), as we have already discussed. For truth in advertising, this misleading term should no longer be used. Instead we should refer to alternation in an honest and transparent manner. It is not randomization in any sense of the word. No partial credit is due for going through the motions with the old college try. Any given trial is properly randomized, or is not. There really is no in between, and if alternation is used, then the trial is not properly randomized and its results cannot be trusted. This remains true, by the way, even if an element of randomization is used to select between the two allocations sequences ABABABAB … and BABABABA … Speaking about element of randomization such as blocks we should mention that it cannot guarantee the comparability of groups and, therefore, internal validity of the study remains questionable.

Minimization & adaptive procedures

Even among the enlightened group of researchers who recognize the folly in using alternation, there still seems to be overwhelming tolerance for minimization (and other similar adaptive allocation procedures), which is not generally equated with alternation or cast under the umbrella of quasi-randomization. In fact, even so prominent a research group as the Cochrane Collaboration explicitly exempts minimization from the perils of being considered less than true randomization. The

risk of bias assessment tool [10] states that "minimization may be implemented without a random element, and this is considered to be equivalent to being random."

Let us examine this statement. In fact minimization is not one single allocation procedure, but, rather, represents an entire family of procedures, indexed by the allocation probabilities when there is a differential imbalance caused by allocating a given patient to one treatment group relative to the other. This may be implemented with biasing probabilities, which themselves may or may not depend on how large the difference in imbalance would be. But the extreme form, indeed the one mentioned [10], does not use biasing probabilities at all. Instead, the allocation is deterministic, to whichever treatment group will minimize the imbalance (hence the name).

True, there is still randomization in case of ties, and this includes the first patient to be allocated, so there necessarily is *some* randomization, even with the pure form of minimization. But since there is a very real possibility that the first allocation is the only one to be decided with any element of chance, we are back in essentially the same situation as alternation augmented with a random element to choose one of the two allocation sequences. This is not enough. Just as with alternation, so too is it the case that minimization is predictable, although some explanation is warranted here.

It is true that one cannot predict, prior to identifying a patient, which treatment will come up next. In this sense, there does seem to be allocation concealment with minimization. However, this changes once the patient to be put on trial is identified. Once we have the patient in hand, we know the age, weight, ECOG score or whatever variables are used to define the imbalance function. We are therefore in a position to calculate the imbalance resulting from placing this patient in either group, so we can certainly determine where this patient would go if enrolled, even if a different patient might, at this same juncture, end up in the other treatment group. In other words, minimization does offer unconditional allocation concealment, but not conditional allocation concealment [11]. Does this matter?

In fact it does matter. Even with unconditional allocation concealment intact, a violation of conditional allocation concealment still leads to profound vulnerability to the same type of selection bias that plagues trials without unconditional allocation concealment. That is to say, an investigator can still deny enrollment to patients who would not be convenient to the desired outcome. Let us posit the existence of other indicators, not included in the minimization imbalance function, that predict the ability of the patient to respond well to treatment. Without implicating all, or even most, investigators, we note that we cannot rely on the honesty and general good intentions of all investigators when evaluating a system. Therefore, when designing a system, we must impose a stress test based upon a worst-case scenario. Along these lines, we consider how much damage an investigator can do, if so inclined, when minimization is used in its pure form. This is a relevant consideration, by the way, even in the absence of any evidence that any investigators ever would do harm.

Minimization allows an investigator to determine, for each patient screened, the treatment group to which this patient would be allocated if enrolled. If the investigator wants to ensure that one treatment group looks better than the other, then he or she can accept 'good' responders (as defined by the key variable not considered by the imbalance function) for the favored treatment group and 'bad' responders for the other treatment group, while denying enrollment to 'bad' responders who would have ended up in the favored treatment group and 'good' responders who would have ended up in the other treatment group. The susceptibility itself is ample reason to avoid using minimization [12], [13], whether or not there is any evidence that the potential for harm has ever actually been realized.

But in fact this potential has been realized, rather often in fact. Chapter 3 of [6] lists 30 trials (not all of them using minimization) in which this type of selection bias is strongly suspected, and Fayers and King [14] discussed another. Brown *et al.* [15] found that roughly one in six investigators have admitted to tracking past allocations so as to predict future ones. How many more do so without admitting it? Strong encryption is needed to curb this type of selection bias resulting from a lack of allocation concealment, and minimization simply cannot offer strong encryption. Therefore it is not in any way, shape or form equivalent to true randomization, and should never be treated as such.

The attempt to balance treatment arms according to certain factors can lead to imbalance in other factors that were not considered in minimization procedure; moreover, as mentioned before, minimization does not prevent allocation prediction. Hence, the contrary position of the Cochrane Collaboration should meet with as much approval as a proposal that for certain endpoints response rates below 20% will be treated as equivalent to response rates above 20%, or that patients experiencing a certain adverse reaction will be considered to not experience this adverse reaction. This is a bewildering statement that takes us collectively further away from reality, and therefore it represents a grave disservice to not only the integrity of the science itself but also the patients who rely on researchers to conduct medical research in such a way that the results best approximate the truth. Clearly, then, we need truth and reality, as opposed to arbitrary decisions that violate truth.

The permuted blocks design

The PBD is the most common approach for randomization [16 1718]. Presumably this is because it helps to avoid chronological bias [5], although it is unlikely that this basis for its use is recognized by very many of the researchers who apply it. It seems far more likely that they use it only because they were told to, or because they see everyone else doing it. This is a shame, because a clear understanding of why we use a given method establishes a level playing field upon which other methods may fairly compete. Without this, a method with better performance is up against the legendary status of the established method, and this is a battle that is almost never won. Indeed, the PBD does have legendary status, is often the only randomization method taught in classes and is the only method offered by randomization.com.

This legendary status and monopoly position of the PBD would not be a problem if its efficiency were commensurate with its popularity. But in reality, the PBD is arguably the worst among all true randomization methods; certainly it is one of the worst. It offers almost no encryption, or protection against allocation prediction. Therefore, allocation concealment depends on perfect masking [19], and if this condition is violated or infeasible or even just uncertain (which is to say, always), then selection bias can occur. Since the last allocation of a block is always deterministic [20], the PBD does not prevent selection bias if the block sizes are known [16, 21, 22]. Even if the block sizes are varied and unknown, there is still a possibility of allocation prediction [23], sometimes even more than with fixed block sizes, because investigators can simply guess whichever treatment group so far has fewer patients [6]. Large block sizes are often recommended to reduce predictability [24]. However, this step also increases the vulnerability to chronological bias [4]. Moreover, many trials are stratified, and this necessitates providing balance in even the small strata [25].

The advantage of the PBD, and a possible reason for the high popularity of this method, is it's comparatively simplicity. However, as it was shown by Zhao, the PBD has the highest possibility of selection bias among restricted randomization methods [25]; hence, using the PBD is unwarranted [26]. Taking into account the fact that biased results of clinical trials could lead to detrimental effect on human health [27], simplicity of analysis should never outweigh methodological rigor. Moreover, modern advances in computer technology have almost eliminated the simplicity of analysis as a determinative factor in the selection of the study design [17]. Considering all of the above, the PBD should not be used in clinical trials at all, let alone as a gold standard. If no better methods were known (and indeed the existing better methods do in fact remain unknown to large cohorts of researchers), then the search for better methods should constitute a top research priority [28]. Since better methods have been identified (as we shall discuss in the next section), the priority should shift toward getting these better methods into use and eradicating the PBD from use in practice [29].

MTI procedures

Though the term 'MTI procedure' did not originate until later, the groundwork was laid over 30 years ago when Soares and Wu [2] proposed the big stick procedure. The idea is to ensure comparable group sizes, not only at the end of the trial, but also throughout the trial, much as the permuted blocks procedure does, but using a different approach. Namely, the big stick uses unrestricted randomization, with equal probabilities to each group, until a prespecified MTI is reached, at which point the big stick is invoked to knock the allocation sequence back toward balance. In other words, in this boundary we have a deterministic allocation, with 100% probability of allocating the under-represented treatment.

The big stick controls chronological bias [5] as well as the permuted blocks procedure will. The major distinction is in the encryption they provide, and protection they offer against prediction and selection bias. The big stick, Chen's and the maximal procedures have common features: first, they all provide fewer restrictions than the PBD; second, when there is no imbalance between groups, they use equal allocation to every group (50/50 in case of two treatment groups); but when the MTI is not reached, and imbalance exists, the big stick procedure will apply equal allocation, whereas Chen's procedure will appeal to a specified and fixed biasing probability, and the and maximal procedure will apply conditional biasing probabilities that depend on the extent of imbalance [30]. Therefore, the maximal procedure controls imbalance better than Chen's procedure does.

It is beyond the scope of this work to derive the theory as to why these MTI methods outperform the PBD in terms of resistance to prediction and enhancement of allocation concealment. It suffices to point out that this superiority has already been established [46], and can be understood intuitively by appeal to how each method handles imbalances reaching the MTI. All the procedures we consider force a return toward balance in this case, but only the PBD requires reeling in the imbalance all the way to zero. The MTI procedures differ by forcing only one allocation toward balance. If the MTI is three, and the imbalance reaches this MTI of three, then all procedures force the next allocation to reduce that imbalance to two. The MTI proce-

dures leave it at that. The PBD instead forces the one after that to be one and the one after that to be zero. This represents excessive restriction on the allocation sequence and greater opportunity for investigators to predict upcoming allocations.

Mixing & matching

The eternal quest for the Holy Grail did not end with Ponce de Leon, and in fact it has made its way to research methodology in general, and randomization methods in particular, in which optimal procedures are often sought. See, for example, [31]. Though optimality makes good sense in other contexts, it does not make sense in the context of randomization methods. This is because the objective of randomization is to create comparable comparison groups, and one of the threats to this is the ability of the investigator to predict future allocations. This ability is enhanced if the method to be used is known in advance, which would be the case if one method were identified as being optimal. Just as a football team cannot afford to always pass on third down, and a pitcher cannot always throw a fast ball, so too is it the case that for the very same reason we cannot always use one randomization method. Combining different randomization procedures helps to control both chronological and selection biases better than using just one technique [32], at least if the various methods are used together in a strategic manner. As noted, there is trade-off between chronological and selection bias for any given randomization procedure [6]. Therefore, it is logical to match procedures that are less susceptible to selection bias with those that better control chronological bias.

For example, investigators might use different MTI values across the different sites of a given trial [32] as a test of robustness of the findings. Concerns about selection bias will be greatly reduced if at least the general trend of the treatment effect appears not only overall but also in those sites (or strata) with the largest MTI values, since larger MTI values tend to reduce the vulnerability to selection bias [4,6]. However, perturbation of group sizes can reduce the study power, so the MTI should never be *too* large in any one site or stratum, and we would also want to see the same trend in the strata with smaller MTI values to rule out the possibility that all we are seeing is an artifact due to chronological bias.

Of course, the variation in randomization methods used in a given trial can be far more comprehensive than varying only the MTI for a single given procedure. One can also use different types of randomization techniques across the strata to vary, and therefore to reduce, the vulnerability to bias [32]. If a chain is only as strong as its weakest link, then we would of course want to limit ourselves to only the best randomization procedures. Therefore, we would not want the permuted blocks procedure or minimization to play any role whatsoever in any undertaking as important as a randomized clinical trial. But, alas, this paradigm goes only so far in accurately describing the problem we face. As noted earlier, it is to our advantage to provide as little information as possible to the investigators regarding how the trial is to be randomized, both overall and specifically at their site. The maximal procedure may be far superior to the permuted blocks procedure, but yet there still may be some value, just to avoid being predictable, in mixing in the use of the permuted blocks procedure (and minimization) in some of the strata.

It is not common to see unrestricted randomization used in actual trials, since this allows for large imbalances in the sizes of the treatment groups. But it might be far more palatable to instead use unrestricted randomization in one stratum, or in a few strata, of a trial that uses other methods in the other strata. Some covariate-adaptive designs [25] combine the big stick procedure or block urn design [33] with biased-coin minimization. Although this method cannot be considered to be a combination of two randomization methods (since minimization is not true randomization [9]), covariate-adaptive design serves as an example of technique mixing in clinical trial design. The combination of different methods used in conjunction will limit the ability of the investigator to predict the type of randomization procedure being used, and, therefore, will also limit the ability of the investigator to predict the next treatment allocation.

However, mixing of randomization procedures have some challenges; first of all it is not easy to select appropriate randomization procedures to be mixed. The choice of randomization procedures depends of many factors, including size of the trial, enrollment protocol, number of treatment arms, stratification etc.

Some trials require perfect treatment balance in each site that limits our ability to combine different randomization techniques. Another example – when stratification cannot be used, in this case randomization procedure can be mixed with minimization. Hence, it is not possible to recommend universal combination of randomization techniques, selection of procedures should be done according to needs each particular study.

Executive summary & conclusion

Randomization is one of the cornerstones of trial quality, but unfortunately it does not receive attention commensurate with its importance. Far too many researchers are content to use permuted blocks randomization, minimization or even alternation, despite the well doc-

umented deficiencies with all of these methods. The MTI methods have been established as far superior in terms of minimizing susceptibility to prediction, and, therefore, they do a much better job of enhancing allocation concealment. Moreover, the MTI methods are easy to use, so there really is no compelling reason to use an inferior randomization method, such as permuted blocks, and there are good reasons not to.

The common fallacy taking place in selection of a randomization technique is to start with sample size. For example, if sample size is big enough, simple randomization can be used. In fact, first of all researcher should think about allocation concealment. Proceeding from this point, simple randomization or permuted blocks cannot be a premier choice, as well as alternation and minimization cannot be a choice at all. Maximum tolerated imbalance methods establish better protection from selection and chronological biases, it is logical to assume their priority. Mixing and matching different randomization methods help to achieve further reduction of prediction. However, there is no overall standard in choosing of exact procedure, since goals and conditions of every trial vary.

There is also a sound rationale for avoiding the pitfall of always using the same randomization method in all trials, because this in and of itself creates vulnerability even if the one method used happens to be optimally resistant to prediction. So, we seek a broader and more inclusive set of randomization procedures even if the maximal procedure seems to be the single best one. The variation we envision would be not only across trials, but also across strata within a given trial, and the variation would encompass both randomization methods and MTI values used with a given method. This mixed approach would provide the best opportunity to build into the structure of the trial the resistance to prediction that is so necessary for internal validity.

Future perspective

There is no doubt that randomization procedures will receive further development. Future elaboration of randomization techniques will be connected with the development of the maximum tolerated imbalance procedures. Moreover, in the process of decision-making about the selection of exact procedure assessment of possible risk of bias will be involved.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

References

1 Berger VW. Conflicts of interest, selective inertia, and research malpractice in randomized clinical trials: an unholy trinity. *Sci. Eng. Ethics* 21(4), 857–874 (2015).

2 Soares JF, Wu CFL. Some restricted randomization rules in sequential designs. *Commun. Stat. Theory Methods* 12(17), 2017–2034 (1983).

3 Chen YP. Biased coin design with imbalance tolerance. *Commun. Stat. Stochastic Models* 15(5), 953–975 (1999).

4 Berger VW, Ivanova A, Deloria-Knoll M. Minimizing predictability while retaining balance through the use of less restrictive randomization procedures. *Stat. Med.* 22(19), 3017–3028 (2003).

5 Matts JP, McHugh RB. Conditional markov chain design for accrual clinical trials. *Biomet. J.* 25, 563–577 (1983).

6 Berger VW. Selection Bias and Covariate Imbalances in Randomized Clinical Trials. John Wiley & Sons, Chichester, UK (2005).

7 Altman DG. The scandal of poor medical research. *BMJ* 308(6924), 283–284 (1994).

8 Ioannidis JPA. Why most published research findings are false. *PLOS Med.* 2(8), e124 (2005).

9 Berger VW, Bears J. When can a clinical trial be called 'randomized'? *Vaccine* 21(5–6), 468–472 (2003).

10 The Cochrane Collaboration's tool for assessing risk of bias. http://ohg.cochrane.org/sites

11 Berger VW. A note on conditional and unconditional allocation concealment. *Clin. Invest.* 4(5), 407–413 (2014).

12 Berger VW. Minimization, by its nature, precludes allocation concealment and invites selection bias. *Contemp. Clin.Trials* 31(5), 406 (2010).

13 Berger VW. Minimization: not all it's cracked up to be. *Clinical Trials* 8(4), 443 (2011).

14 Fayers PM, King M. A highly significant difference in baseline characteristics: the play of chance of evidence of a more selective game? *Qual. Life Res.* 17(9), 1121–1123 (2008).

15 Brown S, Thorpe H, Hawkins K, Brown J. Minimization – reducing predictability for multi-centre trials whilst retaining balance within centre. *Stat. Med.* 24(24), 3715–3727 (2005).

16 Matts JP, Lachin JM. Properties of permuted-block randomization in clinical trials. *Control Clin. Trials* 9(4), 327–344 (1988).

17 Zhao W, Weng Y, Wu Q, Palesch Y. Quantitative comparison of randomization designs in sequential clinical trials based on treatment balance and allocation randomness. *Pharm. Stat.* 11(1), 39–48 (2012).

18 Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomization in leading medical journals: review and reanalysis. *BMJ* 345, e5840 (2012).

19 Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: importance of allocation concealment and masking. *Fam. Med.* 39(2), 132–137 (2007).

20 Berger VW. Do not use blocked randomization. *Headache* 46(2), 343 (2006).

21 Follmann D, Proschan M. The effect of estimation and biasing strategies on selection bias in clinical trials with permuted blocks. *J. Stat. Planning Inference* 39(1), 1–17 (1994).

22 Proschan M. [Influence of selection bias on Type I error rate under random permuted block designs]. *Statistica Sinica* 4, 219–231 (1994).

23 Berger VW. Varying block sizes does not conceal the allocation. *J. Crit. Care* 21(2), 229 (2006).

24 Efird J. Blocked randomization with randomly selected block sizes. *Int. J. Environ. Rec. Public Health* 8(1), 15–20 (2011).

25 Zhao W. A better alternative to stratified permuted block design for subject randomization in clinical trials. *Stat. Med.* 33(30), 5239–5248 (2014).

26 Berger VW. Misguided precedent is not a reason to use permuted blocks. *Headache* 46(7), 1210–1212 (2006).

27 Berger VW, Matthews JR, Grosch EN. On improving research methodology in medical studies. *Stat. Methods Med. Res.* 17(3), 231–242 (2008).

28 Berger VW. Failure to look beyond blocks is a mistake. *Methods Inf. Med.* 54(3), 290 (2015).

29 Berger VW. A tale of two randomization procedures. *Biom. Biostat. Int. J.* 2(3), 00031 (2015).

30 Berger VW, Grant W. A review of randomization methods in clinical trials (chapter 2). In: *Clinical Trial Biostatistics and Biopharmaceutical Applications.* CRC Press, Oxford, UK 47–48 (2015).

31 Hu F, Rosenberger WF. Evaluating response-adaptive randomization procedures for treatment comparisons. *JASA* 98(463), 671–678 (2011).

32 Berger VW, Grant WC, Vazquez LF. Sensitivity designs for preventing bias replication in randomized clinical trials. *Stat. Method Med. Res.*19(4), 415–424 (2010).

33 Zhao W, Weng Y. Block urn design – a new randomization algorithm for sequential trials with two or more treatments and balanced or unbalanced allocation. *Contemp. Clin. Trials* 32(6), 953–961 (2011).